Automated data management

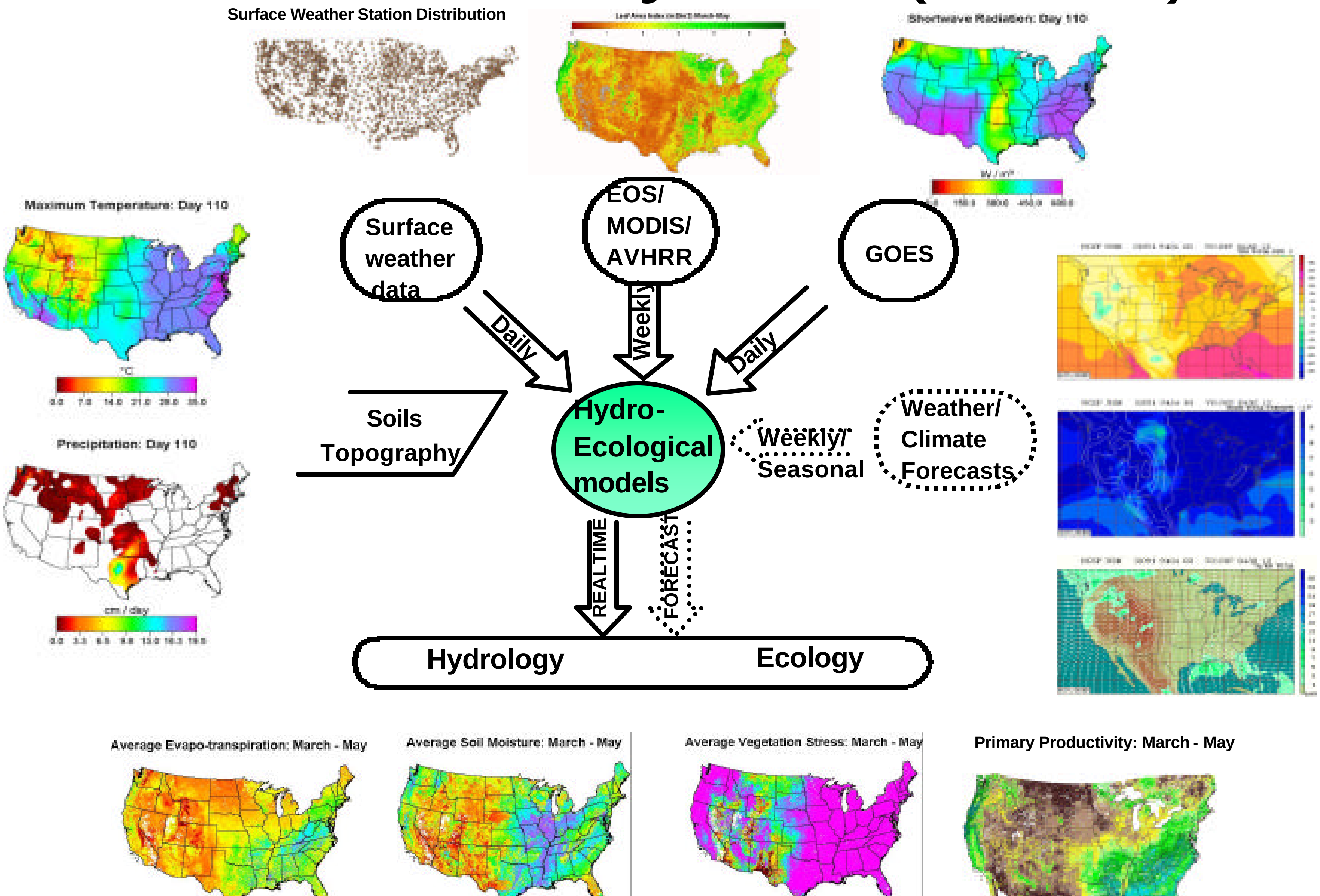
- Principle Investigators
  - Keith Golden, NASA Ames
  - Ramakrishna Nemani, University of Montana, NTSG

Goals

- *Integrate state-of-the-art information technology, weather/climate forecasting, ecosystem modeling, and satellite remote sensing in a single system to enable better management of floods, droughts, forest fires, crop production, and human health*

- Provide an interface in which users merely describe the data products to be generated and programs to generate them are automatically synthesized and executed.

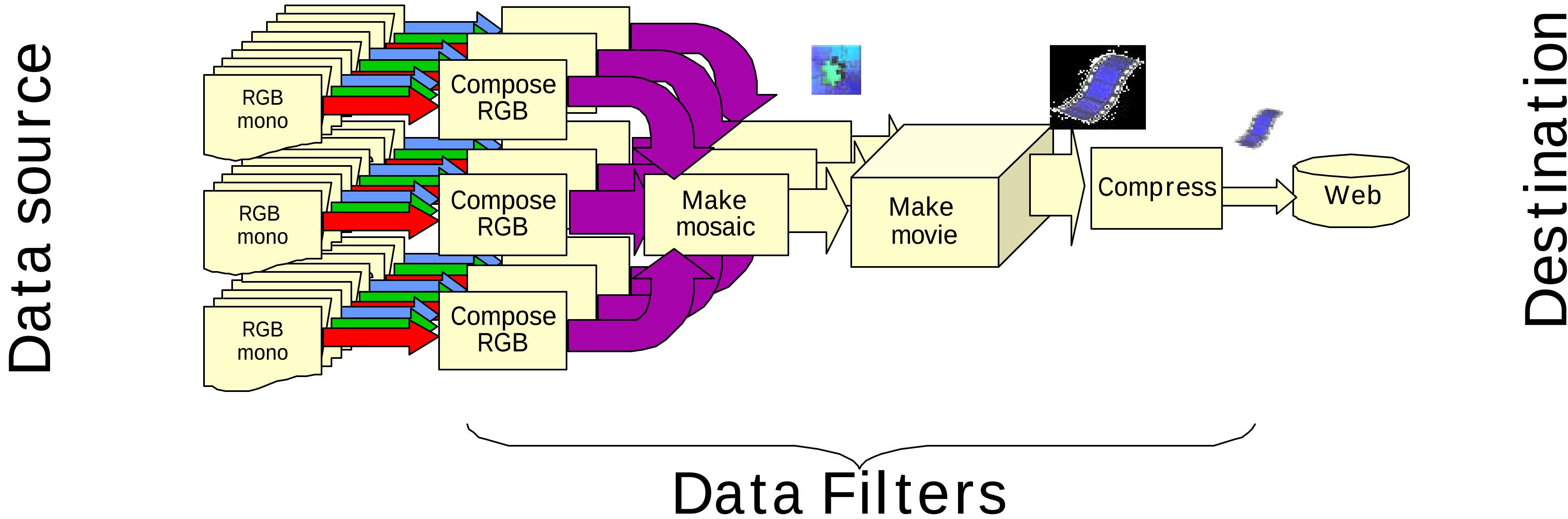
Terrestrial Observation and Prediction System (TOPS)



Technical Problem Statement

- *Create flexible framework that facilitates integration of Earth science data and algorithms*
- Create planner-based system for data-management
  - Support rich representation of data inputs, goals and data-processing programs

Data-flow programs



Output of one command directed to input of another

Data filters: manipulate data

Technical Approach (1)

- Planner-based software robot for data management

- Desired data products given as planner goals

- Raw data available specified in database (“initial state”)

Technical Approach (2)

- Generating data-processing plans requires
  - Very large universes (lots of objects)
  - Creation, destruction, copying of objects
  - Incomplete information

Data Goals and Metadata

- Data product specification
  - What information is contained
  - How information is encoded in data
  - Where the data files are stored/delivered
  - Time that the information pertains to

Planner algorithm

- Heuristic-search constraint-based planner

- Heuristics = distance estimates from primitive graph-based analysis of planning problem

- Entire search space (for fixed-length plans) converted to constraint network

- Planner choices enforced by assigning values to variables in constraint network

Technical Approach (3)

- Framework implementation in Java
  - Flexibility
  - Portability

Technical Approach (4)

- Distributed Architecture
  - Java RMI
  - Java Activatable Framework

Data and NASA Relevance

- Input Data
  - Part of the input comes from NASA's MODIS instrument on Terra satellite platform
- Output Data

Accomplishments (1)

- Action language for data-processing domains

- Prototype of planner that accepts language, generates dataflow plans.

- Preliminary database for storing metadata

Accomplishments (2)

- Designed and implemented prototype of the framework architecture

- Designed and implemented prototype database system that is integrated with the framework

Implemented prototype of XML metadata

Accomplishments (3)

- Implemented prototype Web interface to the database and local production system

- Applied all of the above to NTSGImage project

  - Used the framework architecture together with the database system to designed an automated distributed

Technical Significance (1)

- Demonstrated new approach for managing scientific data by providing goal-directed automation of data processing and generating seachable descriptions of data products.

  - Large data volume, “batch” processing.

  - Many scientific goals/questions

Technical Significance (2)

- Large improvements in flexible development of new Earth science algorithms

- Easier and more flexible production and distribution system design and implementations

  - Mini DAAC

URLs

- http://ic.arc.nasa.gov/people/kgolden/softbots.html
- http://ic.arc.nasa.gov/people/kgolden
- http://www.forestry.umt.edu/ntsg/Projects/TOPS/

Facilities Used

- Work performed at NASA Ames and University of Montana NTSG
- NASA computing facilities limited to personal workstations
- University of Montana Facilities

Personnel Assigned to Projects

- Ames Personnel
  - Dr. Keith Golden (Planner, Language/Parser, Constraint Network)
  - Dr. Wanlin Pang (Constraint Network)

References (1)

- Golden, K., Frank, J. 2002. *Universal Quantification in a Constraint-based Planner*, Proc. of 6th Int'l Conference on AI Planning and Scheduling (AIPS 2002)

- Golden, K. 2001. *A planner-based approach to automated processing and tracking of mission data*. International Symposium on Artificial Intelligence, Robotics and Automation for Space (i-SAIRAS 2001)

- Golden, K. 2000. *Acting on information: a plan language for manipulating data*. In Proceedings 2nd International NASA Workshop on Planning and Scheduling for Space available as NASA Conference

References (2)

- **Nemani**, R.R., M.A. White, P. Votava, J. Glassy, J. Roads and S.W. Running. 2002. *Biospheric forecast system for natural resource management*. Proc. 4th Int'l Conference on Integrating GIS and Environmental modeling (GIS/EM4), Goodchild, M., M. Crane and B. Parks (eds), Banf, Canada.

- Nemani, R.R., M.A. White, P. Thornton, K. Nishida, S. Reddy, J. Jenkins and S.W. Running. 2002. *Recent trends in hydrologic balance have enhanced the terrestrial carbon sink in the United States*. Geophysical Research Letters (May

References (3)

- Nemani, R.R., P. Votava, J. Roads, M. White, P. Thornton and J. Coughlan. 2002. *Terrestrial Observation and Prediction System: Integration of satellite and surface weather observations with ecosystem models*. Proceedings of IGARSS 2002, Toronto, Canada.

- P. Votava, R. Nemani, C. Bowker, A. Michaelis, A. Neuschwander, J. Coughlan. 2002. *Distributed Application Framework for Earth Science Data Processing*. IEEE Geoscience and Remote Sensing Conference proceedings.

Presentations

- A planner-based softbot for data processing
  - AIPS workshop on Real-world Planning: Beyond Operator Sequencing, 2002
- Terrestrial Observation and Prediction System